Measurement Invariance of an Essential Skills Assessment for Students With and Without ACT Testing Accommodations

Yao Sun, Jill McVey, and Cristina Anguiano-Carrasco

Accurate and equitable assessments are key for making valid comparisons and understanding the impact of essential skills interventions across diverse student populations, including students with disabilities (Klingner & Edwards, 2006). Measurement invariance, or whether an assessment measures the same construct in the same manner across different groups, is crucial for fairness in educational evaluations and for preventing biased conclusions (Kline, 2013).

This study was completed using data from ACT's essential skills assessment for high school students (ACT, 2024), which measures five essential skills: Sustaining Effort, Getting Along with Others, Maintaining Composure, Keeping an Open Mind, and Social Connection. Each skill aligns one-to-one with the Big Five personality framework (Walton et al., 2023). Our study tested measurement invariance for a subset of items between two groups: students with and without testing accommodations on the ACT®, which are given in line with the Americans With Disabilities Act. Students receiving accommodations for disabilities include students with neurodevelopmental disabilities (e.g., learning disabilities, ASD), physical-sensory disabilities (e.g., visual impairment, deafness), and/or psychological disabilities (e.g., anxiety, depression). Some examples of accommodations that students may receive include testing in a quiet environment or having extra time. A measurement invariance analysis assesses whether a tool measures the same constructs across groups equivalently (i.e., items show the same psychometric properties regardless of group) and whether any observed differences in scores reflect true differences rather than inconsistencies in measurement.

Method

Participants

The sample consisted of high school students in grades 9 through 12 who completed the essential skills assessment between August 2020 and January 2021. Participants with and without testing accommodations were matched based on demographic characteristics including gender, race/ethnicity, and grade level. The matched dataset contains data from 477 students without testing accommodations and 477 students with testing accommodations. Among the students without accommodations, 207 identified as female, 266 as male, and 3 as another gender; 1 student did not report their gender. In terms of race and ethnicity, 334 students identified as White, 58 as Hispanic or Latino/a, 32 as Black or African American, 15 as bi/multiracial, 6 as Asian, 14 as American Indian or Alaska Native, and 2 as Native Hawaiian or Other Pacific Islander. An additional 16 students did not report their race or ethnicity. The

sample included 262 students in 9th grade, 139 in 10th grade, 57 in 11th grade, and 19 in 12th grade.

The sample of students with testing accommodations had a similar demographic composition. Among them, 207 identified as female, 264 as male, and 2 as another gender; 4 students did not report their gender. Regarding race and ethnicity, 336 students identified as White, 50 as Hispanic or Latino/a, 33 as Black or African American, 15 as bi/multiracial, 2 as Asian, 13 as American Indian or Alaska Native, and 1 as Native Hawaiian or Other Pacific Islander; 27 students did not report their race or ethnicity. This group included 263 students in 9th grade, 137 in 10th grade, 55 in 11th grade, and 22 in 12th grade.

Materials

Students completed a multi-method online assessment designed to measure five core essential (ES) skills: Sustaining Effort, Getting Along with Others, Maintaining Composure, Keeping an Open Mind, and Social Connection. These skills are aligned with the Big Five personality traits and are assessed using three item formats: Likert-type items, situational judgment tests (SJTs), and forced choice (FC) items. Likert items ask students to rate how much they agree with specific statements, while SJTs present brief scenarios followed by possible responses that students rate based on how likely they are to act that way. In contrast, FC items present sets of equally positive statements and ask students to choose which statement is most and least like them. While this format helps reduce social desirability and comparison to others, it reflects students' preferences for certain traits rather than indicating whether and to what degree students actually possess a given trait. For this reason, FC items do not produce data that can be used to compare groups in a measurement invariance analysis. Therefore, only Likert and SJT items were included in the present analyses. Each ES skill was assessed using eight 6-point Likert-type items. SJTs included two scenarios per skill, each followed by five behavioral response options rated on a 5-point scale.

Data Analysis

Guided by measurement invariance theory (Vandenberg & Lance, 2000), we adopted confirmatory factor analysis (CFA) to determine whether ES assessment items yield equivalent results for students with and without ACT testing accommodations. We conducted a series of increasingly robust measurement invariance analyses in R using the lavaan package (Rosseel, 2012). For each of the five factors measured in the assessment, we tested three levels of measurement invariance: configural, metric, and scalar (Meredith, 1993; Vandenberg & Lance, 2000). Configural invariance refers to the condition in which the same factor structure is present across groups, indicating that the construct is conceptualized similarly. Metric invariance (also known as weak invariance) requires that factor loadings—that is, how strongly each item relates to the underlying construct—be equal across groups, suggesting that items relate to the latent construct in the same way. Scalar invariance (also known as strong invariance) involves equal item intercepts across groups, allowing for meaningful comparisons of latent means. Metric invariance is considered the minimum requirement, while scalar invariance is a more stringent criterion for interpreting scores across groups in the same way (Putnick & Bornstein, 2016).

To begin with, we ran CFA for each factor for all Likert and SJT items and retained only the items that yielded good factor loadings (> .40) for further analyses. The measurement model for each factor yielded satisfactory model fit (CFIs and TLIs > .90, RMSEAs < .08, SRMRs < .06).

Then we tested the configural invariance of each of the five factors to confirm that the factor structure of each one was consistent across groups (students with and without testing accommodations). For each factor, we specified the two groups in the CFA model, then used a chi-square test to compare the model fits before and after specification of the two groups. The results showed that the model fits of all factors did not differ significantly before and after specification of groups (ps > .05), thus supporting configural invariance.

Next we tested the metric invariance of each factor, meaning we tested whether factor loadings were equivalent across groups. Similarly, for all factors, we constrained factor loadings so they were equal across groups and ran a series of chi-square tests to compare the model fits of the CFA models before and after constraining. The results showed that they did not significantly differ (ps > .05), supporting a metric invariance of the factors.

Finally we tested scalar invariance by constraining the intercepts of the items so they were equal for both groups. The results of the chi-square tests showed that the model fits significantly dropped after we constrained the intercepts of the items across groups. This being the case, we ran a series of score tests, or a Lagrange Multiplier test, to explore whether releasing one or more constrained parameters in the model would improve its fit to the data (Bentler & Chou, 1992). The score tests also allowed us to detect items that were potentially causing non-invariance in the measurement.

Results

The model fit indices of each step for each factor are summarized in Table 1. All the skills except Getting Along with Others showed significant model fit decreases when intercepts were constrained. In other words, we were able to establish full scalar invariance only for this subset of the Getting Along with Others factor. For the other four factors, one to three problematic items were identified (summarized and presented in Figure 1). After excluding items that were causing non-invariance, we observed partial scalar invariance for these four factors.

An interpretation of the results is that while this subset of items generally supports reliable mean comparisons between the with-accommodations and without-accommodations groups, some items exhibited differing intercepts between the groups, indicating that students with and without testing accommodations start from different points on the scale being measured. In other words, for the same latent level of factor (for example, Social Connection), students with and without accommodations reported being at different skill levels (for example, students without ACT testing accommodations scored higher on an item about motivations to work when they are tired than students with testing accommodations).
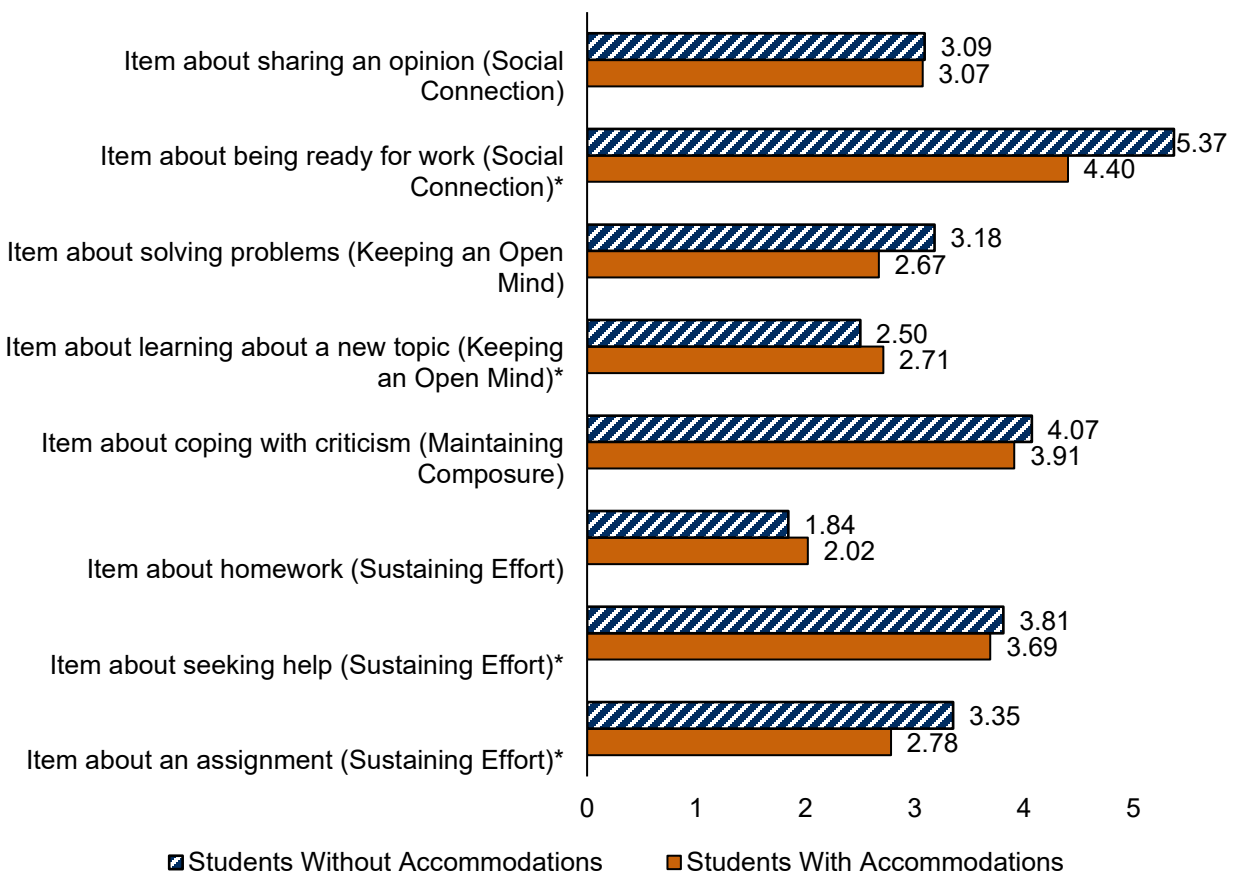
Findings are summarized as follows:

- We established configural invariance, which indicates that the five factors being measured are understood similarly by students in each group, thus supporting the validity of the assessment's factor structure for this subset of items.

- By confirming metric invariance, we ensured that responses can be meaningfully compared across groups, as the association between items and the latent constructs is consistent.

- We established only partial scalar invariance, which suggests that this subset of items is effective for overall group comparisons but indicates potential biases introduced by some items, specifically with the intercepts of the items.

- The final model included 9 items for the Getting Along with Others factor, 8 items for the Keeping an Open Mind factor, 6 items for the Maintaining Composure factor, 10 items for the Social Connection factor, and 14 items for the Sustaining Effort factor, supporting cross-group validity for a version of the assessment that is suitable for comparing the scores of students with and without testing accommodations.

In conclusion, our measurement invariance analysis provides evidence for the overall feasibility of cross-group comparisons using this subset of items as it pertains to students receiving accommodations. Results also highlight the necessity of careful interpretation due to partial scalar invariance. Practitioners and researchers should be mindful of these limitations when using the tool for cross-group comparisons.

Table 1. Model Fit Indices of Measurement Invariance Models

Category		χ^2(DF)	CFI	TLI	RMSEA	SRMR
Getting Along with Others	Configural	144.26(52)	0.95	0.93	0.06	0.04
	Metric	158.10(60)	0.95	0.93	0.06	0.05
	Scalar	168.84(68)	0.94	0.94	0.06	0.05
Keeping an Open Mind	Configural	110.33(36)	0.96	0.94	0.07	0.04
	Metric	116.03(43)	0.96	0.95	0.06	0.04
	Scalar	151.28(50)***	0.94	0.94	0.07	0.05
	Partial scalar	119.47(48)	0.96	0.95	0.06	0.04
Maintaining Composure	Configural	60.14(16)	0.95	0.91	0.08	0.04
	Metric	66.54(21)	0.95	0.93	0.07	0.04
	Scalar	80.26(26)*	0.94	0.93	0.07	0.05
	Partial scalar	75.19(25)	0.95	0.94	0.07	0.05
Social Connection	Configural	218.82(64)	0.94	0.91	0.07	0.05
	Metric	231.47(73)	0.93	0.92	0.07	0.05
	Scalar	262.35(82)***	0.92	0.92	0.07	0.06
	Partial scalar	242.14(80)	0.93	0.92	0.07	0.05
Sustaining Effort	Configural	398.13(124)	0.93	0.91	0.07	0.04
	Metric	417.81(136)	0.93	0.92	0.07	0.05
	Scalar	449.39(148)**	0.92	0.92	0.07	0.05
	Partial scalar	431.60(145)	0.93	0.92	0.06	0.05

Note. The model fit was significantly different from the model fit in the previous step at the * $p <$ .05, ** $p <$ .01, and *** $p <$ .001 levels.

Figure 1. Item Intercept Comparisons of Students With and Without Testing Accommodations

Item about sharing an opinion (Social Connection) — 3.09 / 3.07
Item about being ready for work (Social Connection)* — 5.37 / 4.40
Item about solving problems (Keeping an Open Mind) — 3.18 / 2.67
Item about learning about a new topic (Keeping an Open Mind)* — 2.50 / 2.71
Item about coping with criticism (Maintaining Composure) — 4.07 / 3.91
Item about homework (Sustaining Effort) — 1.84 / 2.02
Item about seeking help (Sustaining Effort)* — 3.81 / 3.69
Item about an assignment (Sustaining Effort)* — 3.35 / 2.78

☑ Students Without Accommodations   ■ Students With Accommodations

Note. All items used a Likert response scale. Items marked with an asterisk were responses to a SJT item with five response choices ranging from "very unlikely" to "very likely." All other items were individual Likert items with six response choices ranging from "strongly disagree" to "strongly agree."

References

ACT. (2024). *Mosaic™ by ACT®: Social Emotional Learning Assessment technical manual.*
https://www.act.org/content/dam/act/unsecured/documents/mosaic-sel-tech-manual.pdf

Bentler, P. M., & Chou, C.-P. (1992). Some new covariance structure model improvement statistics. *Sociological Methods & Research*, *21*(2), 259–282.
https://doi.org/10.1177/0049124192021002006

Kline, R. B. (2013). Assessing statistical aspects of test fairness with structural equation modelling. *Educational Research and Evaluation*, *19*(2–3), 204–222.
https://doi.org/10.1080/13803611.2013.767624

Klingner, J. K., & Edwards, P. A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly*, *41*(1), 108–117.
https://doi.org/10.1598/RRQ.41.1.6

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.
https://doi.org/10.1177/109442810031002

Walton, K. E., Murano, D., Burrus, J., & Casillas, A. (2023). Multimethod support for using the Big Five framework to organize social and emotional skills. *Assessment*, *30*(1), 144–159.
https://doi.org/10.1177/10731911211045744

ABOUT ACT

ACT is transforming college and career readiness pathways so that everyone can discover and fulfill their potential. Grounded in more than 65 years of research, ACT's learning resources, assessments, research, and work-ready credentials are trusted by students, job seekers, educators, schools, government agencies, and employers in the U.S. and around the world to help people achieve their education and career goals at every stage of life. Visit us at www.act.org.